BlockGCN: Redefine Topology Awareness for Skeleton-Based Action Recognition

UNIVERSITY OF MANNHEIM · 澳門城市大學 Universidade da Cidade de Macau City University of Macau · Carnegie Mellon University · 中国科学院深圳先进技术研究院 SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES · Microsoft · Terminus

Yuxuan Zhou† Xudong Yan† Zhi-Qi Cheng* Yan Yan Qi Dai Xian-Sheng Hua

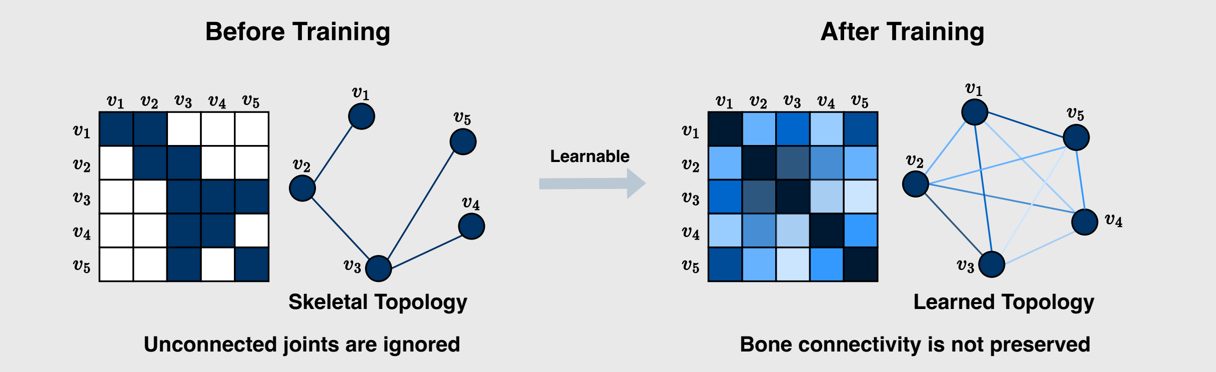
CVPR SEATTLE, WA JUNE 17-21, 2024

INTRODUCTION

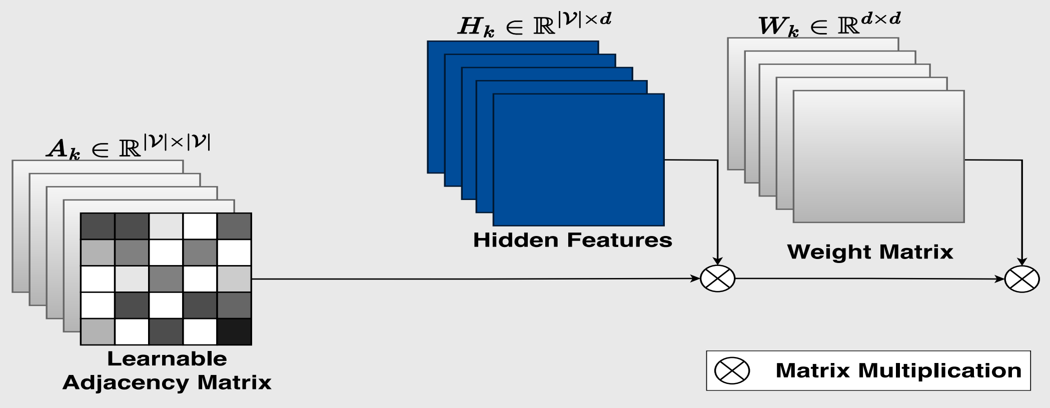
Motivation

We reveal the remaining issues of previous GCNs

Catastrophic Forgetting of skeletal topology



Inefficient multi-relational modeling



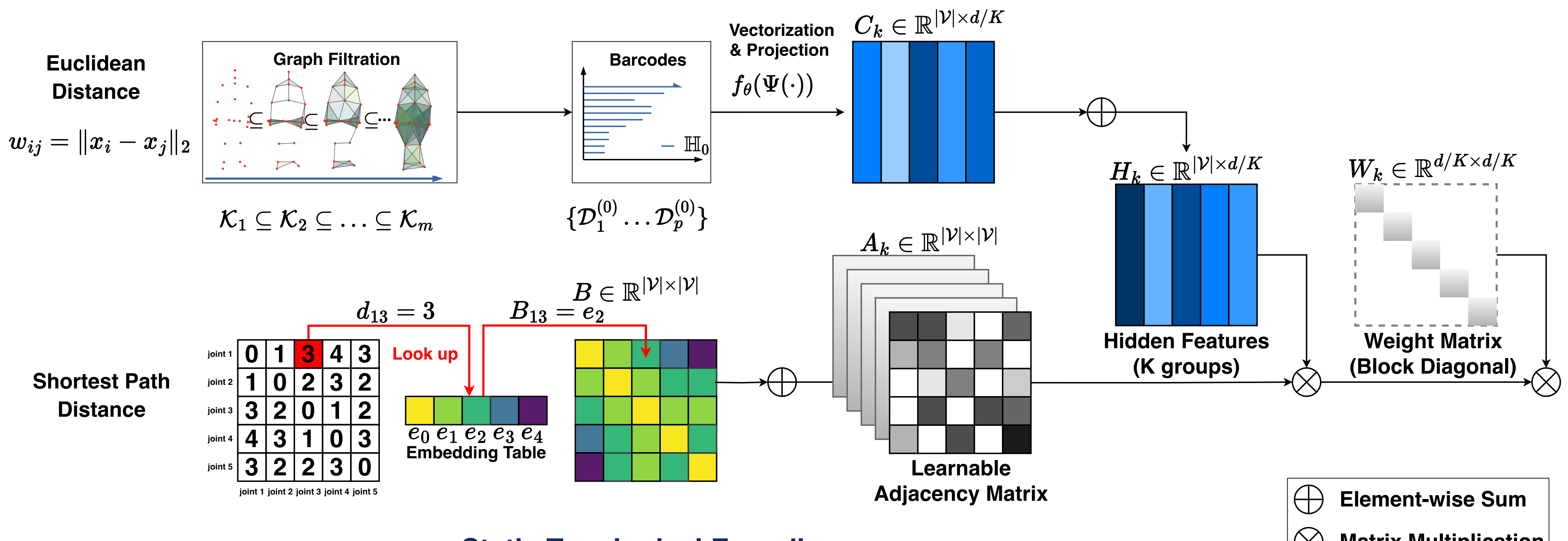
Contribution

- Identifying and restoring the overlooked skeletal topology in advanced GCNs via novel topological encoding schemes.
- Devising BlockGC, an efficient and powerful graph convolutional block.
- Establishing new state-of-the-art performance on standard benchmarks.

† Internship at CMU. Equal contribution.
* Corresponding author.
CodeLink: https://github.com/ZhouYuxuanYX/BlockGCN

METHOD

Dynamic Topological Encoding



Static Topological Encoding

Topological Encoding

Static Topological Encoding: we encode the relative distance between joints on the skeletal graph $\mathcal{G}_S$, using measures like Shortest Path Distance (SPD) or level structure distance.

$$B_{ij} = e_{d_{i,j}} \quad \text{with} \quad d_{i,j} = \min_{P\in Paths(\mathcal{G}_S)} \left\{ |P| \mid P_1 = v_i, P_{|P|} = v_j \right\}$$

Dynamic Topological Encoding: we adopt the differentiable vectorization $\Psi^0 : \{\mathcal{D}_1^0, \mathcal{D}_2^0, \ldots, \mathcal{D}_p^0\} \rightarrow \mathbb{R}^{|\mathcal{V}|\times d'}$ on the barcodes.and project the obtained representation to GCN hidden layers' feature space through a mapping $f_\theta : \mathbb{R}^{|\mathcal{V}|\times d'} \rightarrow \mathbb{R}^{|\mathcal{V}|\times d}$ at each layer:

$$C = f_\theta\left(\Psi^0\left(\mathcal{D}_1^0, \mathcal{D}_2^0, \ldots, \mathcal{D}_p^0\right)\right)$$

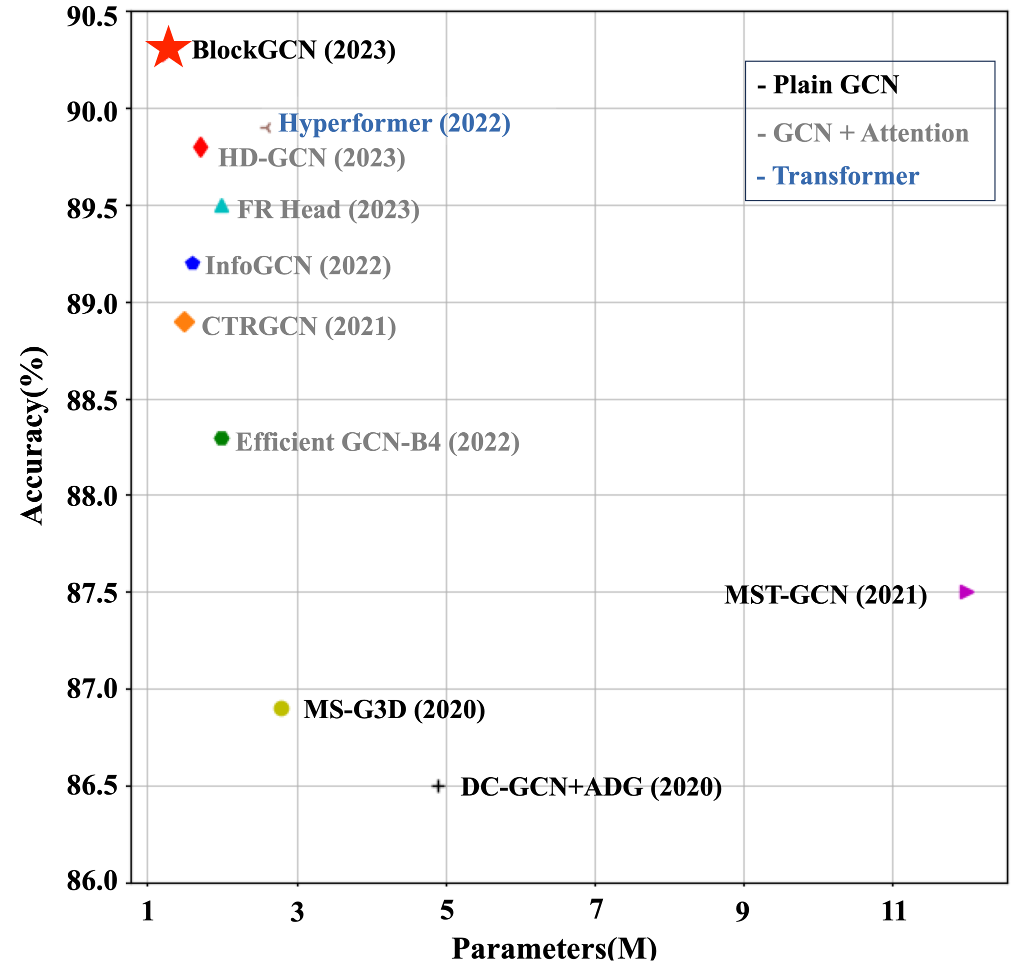
Efficient Multi-Relational Modeling

We propose BlockGC, which efficiently models high-level semantics, reducing computation and parameters while outperforming previous methods. The feature dimension is divided into K groups, with spatial aggregation and feature projection applied in parallel within each group. The formula is as follows:

$$H^{(l)} = \sigma\left(\begin{bmatrix}(A_1+B_1)(H_1^{(l-1)}+C_1^{(l-1)}) \\ \vdots \\ (A_K+B_K)(H_K^{(l-1)}+C_K^{(l-1)})\end{bmatrix}\begin{bmatrix}W_1^{(l)} & & \\ & \ddots & \\ & & W_K^{(l)}\end{bmatrix}\right)$$

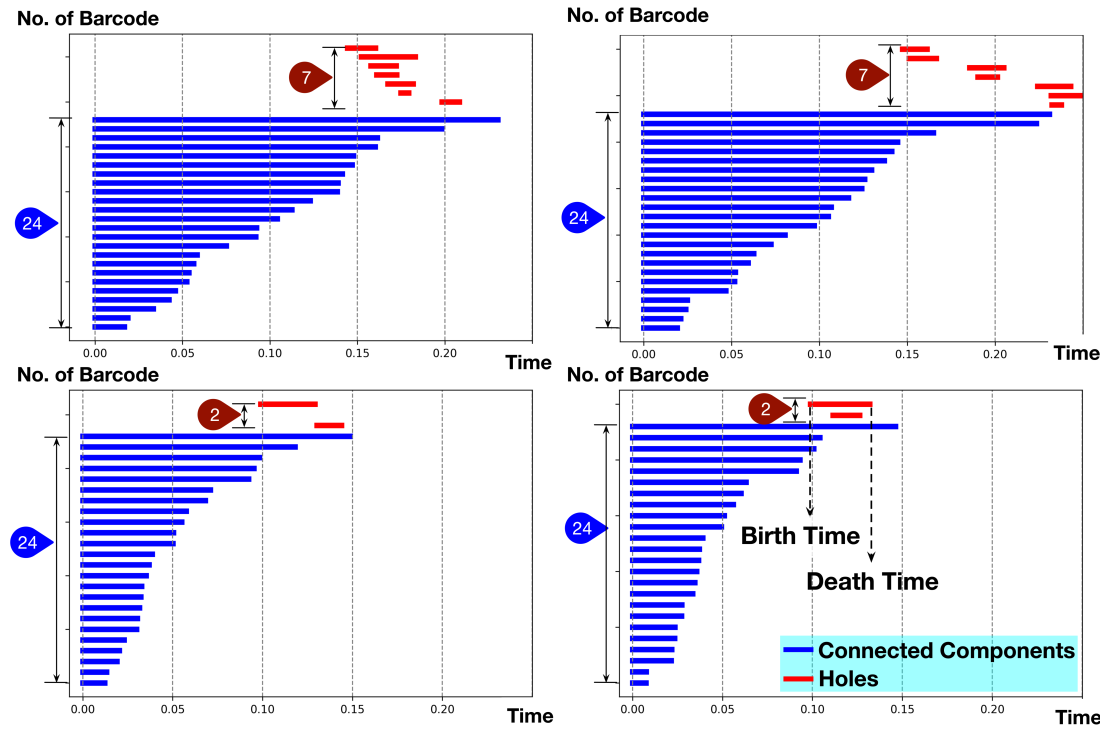
RESULTS

Performance vs. Model Size



Our BlockGCN improves over previous methods w.r.t. both performance and efficiency.

Visualization of barcodes



Barcodes of "brush hair" (top) and "shake hands" (bottom).